https://doi.org/10.33472/AFJBS.6.6.2024.1791-1807

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Modernising Medical Records: Region-based Convolutional Recurrent Neural Network and Connectionist temporal classification-Based Doctor's Handwriting Recognition

Dr. Jaishree Jain[1*], Garima Saroj[2], Ayushi Gautam[3], Bhavya Agrawal[4], Sarthak Gupta[5], Yogendra Narayan Prajapati[6]

[1]Asst. Professor, CSE Department, AKGEC Ghaziabad, India
[2]CSE Department, AKGEC Ghaziabad, India
[3]CSE Department, AKGEC Ghaziabad, India
[4]CSE Department, AKGEC Ghaziabad, India
[5]CSE Department, AKGEC Ghaziabad, India
[6]Asst. Professor, CSE Department, AKGEC Ghaziabad, India

Email: [1]jainjaishree@akgec.ac.in, [2]sarojgarimal609@gmail.com, [3]ayushigtm4@gmail.com, [4]bhagrawal24@gmail.com, [5]sarthakgupta0212@gmail.com, [6]ynp1581@gmail.com

Article Info

Volume 6, Issue 6, May 2024

Received: 08 April 2024

Accepted: 03 May 2024

Published: 28 May 2024

doi: 10.33472/AFJBS.6.6.2024.1791-1807

ABSTRACT:

Handwriting is the way to convey an idea or information through written means. But over the years, due to fewer doctors per population ratio, doctors have become well-known for their illegible cursive handwriting and have become well accepted. The legibility issue of handwritten medical documents, particularly those created by physicians, has long been a significant problem in healthcare. This study presents Doctor's Handwriting Recognition, an innovative solution to tackle the problem of illegible doctor's handwriting in medical records. Our ideation surpasses its function as a recognition system, serving as evidence of technology's ability to unite tradition and innovation in healthcare documentation. Digitising medical records is essential for improving patient care, optimising operations, and safeguarding data. The recognition system uses a Region-based deep Region-based Convolutional Neural network (R-CRNN) that is enhanced with the Connectionist Temporal Categorical (CTC) loss function. This allows the system to adapt to the unique handwriting of individual doctors. Doctor's Handwriting Recognition has the potential to revolutionise healthcare professionals' interactions with handwritten medical information. It offers increased efficiency, enhanced patient safety, and decreased medical errors. Adopting this technological advancement improves healthcare documentation and enhances the accessibility of medical records, ultimately benefiting patient well-being.

Keywords: Doctor's Handwriting Recognition, Region- based Convolutional Recurrent Neural Network (R-CRNN), Connectionist Temporal Categorical (CTC) loss, Deep Learning (DL), Image Segmentation.

1. Introduction

The art of handwriting serves as a means to articulate thoughts and ideas, with doctors being notably renowned for their often indecipherable cursive script—a phenomenon widely acknowledged. This challenge extends beyond medical practitioners, affecting pharmacists tasked with interpreting prescriptions, resulting in a concerning number of medical errors. A documented case in Texas exemplifies the dire consequences, where a cardiologist's prescription for 10 mg Plendil was misread as 20 mg, leading to a tragic patient fatality [1][2]. The intricacies of recognizing human handwriting have spurred technological advancements, particularly in the development of handwriting recognition systems capable of discerning characters across various mediums. Despite existing research in this domain, the challenge of recognizing cursive characters persists, owing to factors such as deformations, size variations, diverse styles, incomplete strokes, flow and Ligatures.

This research endeavours to contribute by crafting a model utilising Deep Region-based Convolutional Recurrent Neural Networks (R-CRNN), specifically tailored for identifying words and numbers within the context of doctors' cursive handwriting. The overarching objective is to mitigate the challenges associated with interpreting illegible handwriting, catering to both medical professionals and individuals without a medical background.

Objective of the Research

Individuals exhibit distinct writing styles, presenting challenges in comprehension, particularly when it comes to doctors who are frequently associated with illegible handwriting [2]. According to Reader's Digest, the primary reason for doctors' messy handwriting lies in the limited time available for patient consultations, forcing them to hurriedly write prescriptions. The added complexity arises from similar-looking medicine names and confusing abbreviations, intensifying the struggle to interpret doctors' handwriting [2][4][5].

Recognition systems play a pivotal role in categorising input patterns into respective entities. Character recognition systems, tailored for character classification, exhibit variations based on the type of characters they handle, be it printed, typewritten, or handwritten characters [6]. While recognizing an individual character is relatively straightforward, complexities arise when dealing with cursive or mixed cursive words. Distinctive writing styles, alphabet shapes, and sizes contribute to the intricate nature of character recognition [7].

Character recognition (CR) assumes critical significance in pattern recognition and image processing. The fundamental objective is to convert intelligible characters into a machine-readable format [8]. Through CR, text embedded in an im- age transforms machine-readable text, tackling the inherent challenges posed by diverse handwritten styles. This process proves particularly crucial in healthcare documentation, where precision and clarity are imperative for ensuring patient safety and facilitating effective communication.

2. Literature Review

Recent explorations into deep learning for handwriting recognition have gravitated towards the integration of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), terminating in the Region-based Convolutional Recurrent Neural Network (R-CRNN) model. This hybrid approach leverages the spatial feature extraction capabilities of CNNs with the sequential data processing prowess of RNNs, making it exceptionally suited for recognizing the nuanced and variable nature of handwritten text [19]. Moreover, the Connectionist

Temporal Classification (CTC) loss function plays a pivotal role in this architecture by enabling the model to handle variable-length input sequences without explicit segmentation, thereby streamlining the recognition process [17].

The application of R-CRNN models, combined with the CTC loss function, has demonstrated promising outcomes in the domain of medical handwriting recognition. Yin et al. (2021) underscored the efficacy of this approach in enhancing the accuracy and efficiency of converting manuscripts into electronic formats. This transition is not only pivotal for ensuring the legibility of prescriptions and medical records but also for fostering a digital healthcare ecosystem that is both robust and accessible.

Furthermore, the incorporation of language models and specialised datasets encompassing medical terminology and drug names has been instrumental in refining the recognition accuracy. By training models on healthcare-specific datasets, the system's proficiency in identifying medical shorthand and jargon has seen considerable improvement [15]. Such targeted training approaches underscore the importance of domain-specific knowledge in enhancing the performance of handwriting recognition systems within the healthcare sector.
Despite the advances, challenges remain, particularly in the aspects of model generalisation and the handling of extreme variations in handwriting styles. Research by Fajardo et al. (2019) highlights the ongoing efforts to tailor deep learning models for greater adaptability and accuracy in real-world applications. These endeavours are crucial for the broader adoption of technology in healthcare settings, where the stakes are inherently high.

The integration of deep learning models for doctor's hand- writing recognition signifies a leap towards mitigating the risks associated with handwritten medical records. It exemplifies the convergence of artificial intelligence and healthcare, aiming to enhance the quality of patient care through improved data accuracy and availability. As this field continues to evolve, future research will undoubtedly focus on addressing the current limitations and exploring innovative applications of deep learning technologies in healthcare documentation.

Existing Studies on Doctor Handwriting Recognition

The study published in South African Family Practice investigates the legibility of doctors' handwriting and its impact on prescription accuracy at the National District Hospital, Bloemfontein. It was found that pharmacists, who are essential in dispensing prescriptions, had the lowest accuracy in reading them. Furthermore, the study highlighted critical reading mistakes that could potentially be lethal and noted that many prescriptions did not meet legal requirements. This research emphasises the importance of addressing the issue of illegible handwriting in healthcare to enhance patient safety and care.

S. Tabassum et al, Deciphering a doctor's prescription can often be a challenge due to their handwriting. In this study, neural network methodologies including CNN and Bi-LSTM were employed to predict doctors' handwriting from medical prescriptions. To aid in normalisation, the CTC loss function was utilised. The model was trained on the IAM dataset, with image acquisition and data augmentation employed for preprocessing. Additionally, the input underwent processing through 7 convolution layers of a neural network. The training phase consisted of 32 epochs, requiring six hours to complete, with loss values depicted on a graph [10].

E. Hassan, H. Tarek, M. Hazem, S. Bahnacy, L. Shaheen and W. H. Elashmawi., A CRNN framework was developed using Python to interpret handwritten English prescriptions and convert them into digital text. The utilised datasets encompassed 66 distinct classes, comprising alphanumeric characters, punctuation marks, and spaces. Given that prescriptions typically consist of two or three words, training was conducted using short texts. Both regular handwriting samples and doctors' prescriptions were utilised for model training. The system achieved a remarkable accuracy rate of 98% after factoring in training duration and data input. The paper emphasised the necessity for further exploration of input-handling techniques to bolster results [11].

L. J. Fajardo et al, The method recognizes handwritten Malayalam letters in cursive using the Hidden Markov Model (HMM). The approach here helps prevent errors due to noise in the scanned image by using a median filter. Additionally, an Artificial Neural Network (ANN) finds the best matching class for input and helps acquire better classification. To make the recognition procedure less complicated, high-quality samples are employed. The speed and precision of the outcomes are higher with this method. Consequently, it is possible to identify the fusion of Malayalam and English characters as future work [12].

K. Gaurav, Bhatia P. K. Et al, This research examines the numerous pre-processing methods used in character identification for a variety of image types, including handwritten forms and papers with complex, multicoloured backgrounds and varying intensities. Various preprocessing methods, including skew detection and correction, contrast stretching, binarization, image enhancement techniques, noise removal, normalisation, and segmentation, as well as morphological processing methods, are covered in this. It was determined that we are unable to fully process the image with a single preprocessing approach. Nevertheless, complete accuracy in a preprocessing system might not be achievable even after using all of the aforementioned strategies [13].

Salvador Espana-Boquera et al., This work proposes a hybrid Hidden Markov Model (HMM) model for offline handwritten word recognition under unconstrained conditions. This uses a Multilayer Perceptron to estimate the emission probability while using Markov chains to simulate the optical model's structural portion [14].

3. Materials and Method

In our system, we exclusively utilised predefined datasets to train the machine learning model for recognizing doctors' handwriting. Python, along with libraries like OpenCV, PIL, and NumPy, were employed for preprocessing the images. The preprocessing steps encompass image resizing, normal- ization, and noise removal to optimise the images for machine learning algorithms. For recognizing doctors' handwriting, our approach leverages the R-CRNN algorithm. R-CRNN, a sophisticated deep learning algorithm, amalgamates CNNS and RNNs to extract features from images and discern patterns in text sequences. The selected datasets were utilised to train the R-CRNN algorithm. Throughout the training phase, the preprocessed images were inputted into the algorithm, allowing it to learn the distinctive handwriting patterns of doctors. As a result, we have established a recognition system that is both comprehensive and proficient in its ability to accurately discern and understand the nuances inherent in doctors' handwriting.

Explanation of the Approach Used

The construction of a strong quantitative framework was guided by valuable insights obtained from the literature study. The decision to use deep learning, namely R-CRNN, was based on its established effectiveness in tasks involving image and sequence recognition. By using the

CTC loss function, the system can effectively account for the temporal aspects of handwriting and so adapt to the distinctive traits exhibited by each doctor and can handle variable length handwriting too. This approach is suitable for the complex task of recognizing handwriting in medical records, providing a well-balanced and novel solution.

Description of the Data Collection Process

The process of data collecting entailed working along with healthcare organisations to get authorised samples of medical records that had been stripped of all identifying information. To promote diversity, the dataset was expanded by including samples from other areas and demographics. The pre-processing stages encompass image normalisation, scaling, and extraction of pertinent features. Thorough quality inspections were carried out to remove any anomalies and protect the integrity of the dataset.

Overview of the dataset used: For this particular study, we have incorporated the IAM Handwritten Forms dataset consisting of a varied assortment of handwritten records, providing a wide range of handwritten samples that accurately represent medical records. The inclusion of this feature ensures that the model can easily adapt to different doctors' handwriting styles, preventing any issues with missing or incorrect letters that may occur due to faster writing. Furthermore, we integrated a comprehensive collection of medicines and medical terminology. This additional dataset contains commonly used shorter abbreviations that correspond to the original names. It enhances the system's ability to recognize a wide range of medical terminology and abbreviations commonly found in handwritten records. The dataset was meticulously selected to include a wide range of handwriting intricacies, ensuring a strong basis for training the recognition engine.

Additionally, for the fine-tuning phase, we acquired diverse pharmacy samples and meticulously annotated them using the VGG Image Annotator (VIA). The VGG annotator facilitated precise labelling and enriched our dataset with over 2000 samples, contributing to the authenticity and diversity of real doctors' handwriting in the training data. The VIA stands as an open-source tool crafted by the Visual Geometry Group (VGG) at the University of Oxford. Under the BSD-2 clause licence, VIA enables manual annotation of images, audio, and video. This tool excels in defining regions within an image and generating textual descriptions for those regions. VIA outputs labels in CSV or JSON format, and its compatibility with Roboflow allows seamless importation and conversion to various object detection formats.

Procedures employed for collecting and preparing the data: In collecting data for the Doctor's Handwriting Recognition project, we meticulously sourced handwritten prescriptions and medical notes from pharmacies, medical facilities, and datasets like the IAM forms dataset. Each sample under- went thorough annotation, including prescription details and additional context. Preprocessing ensures data consistency, and partitioning into training, validation, and testing sets facilitated model evaluation. Augmentation techniques were applied to enhance dataset diversity, and rigorous quality checks were conducted to ensure reliability. These steps ensured a robust dataset for our research study.

Figure 1 outlines the systematic workflow of our data processing system. The initial stage involves the collection of handwritten notes, including a diverse array of doctors' prescriptions, which are meticulously annotated using the VGG Image Annotator for precise labelling. This annotated corpus then constitutes our primary dataset, prepped for the subsequent processing stages.

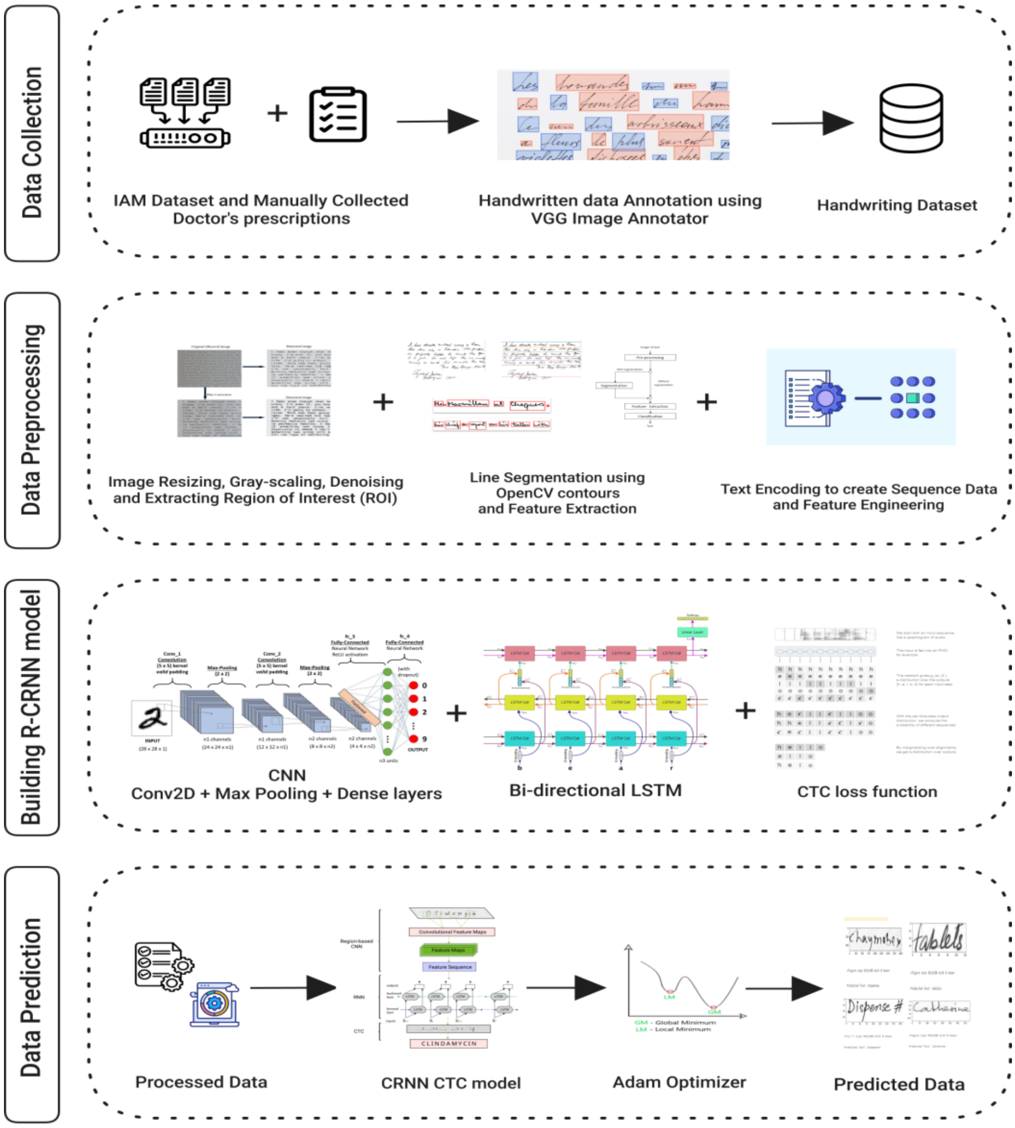

Figure 1. R-CRNN Architectural Schematic for Handwriting Decipherment: Encompassing initial data preparation, convolutional neural processing, bi-directional LSTM for temporal dynamics, Optimization using Adam Optimizer and CTC decoding for final text prediction.

A. Overview of the Techniques and Algorithms Employed for Handwriting Recognition

Explanation of the chosen techniques and algorithms: The research utilises an R-CRNN network combined with the CTC loss function. This combination is pivotal for recognizing variable handwriting styles, as R-CRNN excels in extracting sequential and spatial features from images, while CTC aids in aligning the input sequences with their labels without predefined segmentation.

Data Preparation and Standardization:

In the initial stages of the project, emphasis was placed on data preparation to ensure a solid foundation for model training. The dataset, sourced from a CSV file, underwent meticulous cleaning processes, including the elimination of missing values and the filtration of identities to a character length not exceeding ten. This preliminary step was critical for ensuring dataset consistency and managing computational complexity effectively. Additionally, the creation of dictionaries for character-to-label mapping was pivotal. It enabled the transformation of textual data into a numerical format thereby rendering it interpretable by the model. This foundational work was essential in setting the stage for the advanced stages of model development, underpinning the entire project with a robust dataset optimised for machine learning.

Image Preprocessing and Data Augmentation: Subsequent efforts were directed towards image preprocessing and augmentation, facilitated by the development of a bespoke 'Data Generator' class. This class was instrumental in batch- processing the data, thereby enhancing memory management during the model training phase. It enabled the application of crucial preprocessing techniques, including image resizing and normalisation of pixel values, which were vital for preparing the images for the model. This preparation was crucial in augmenting the model's learning efficacy. Furthermore, the augmentation process introduced data variability, significantly bolstering the model's ability to generalise. This phase of the project underscored the planning and execution of preprocessing strategies, significantly contributing to the model's robustness.

Building the Model: The R-CRNN model architecture was the linchpin of this project. It captured the essence of the preparatory work through the integration of convolutional layers for feature extraction and bidirectional LSTM layers for capturing sequence dependencies. The strategic inclusion of a CTC layer facilitated the recognition of sequences without necessitating explicit segmentation, thereby streamlining the learning process. Upon compilation with the Adam optimizer, the model progressed to the training phase, utilising the prepared data generator for both training and validation purposes. The training phase yielded promising results, with subsequent evaluation and prediction stages showcasing the model's proficiency in translating handwritten text into a legible, human-readable format. This comprehensive methodology, spanning from data preparation to model evaluation, highlighted the project's success in leveraging deep learning technologies to tackle the intricate task of handwriting recognition, thus exemplifying academic rigour and methodological robustness in computational research.

Mathematical Foundations of the R-CRNN Approach: In our approach, the R-CRNN architecture plays a pivotal role in interpreting complex handwriting patterns. R-CRNN combines the spatial feature extraction capabilities of CNNs with the sequential data processing strength of RNNs, specifically employing LSTM units for better handling temporal dependencies. The CTC loss function is utilised to train the network efficiently on unsegmented sequence data.

1. CNN: The CNN layers are designed to automatically and adaptively learn spatial hierarchies of features from input images. A convolutional layer's output F , for a given input image X, can be defined as:

$$F^k = \sigma \left(\Sigma_m \Sigma_n W^k_{mn} \quad X(i+m)(j+n) + b^k\right)$$

where W is the weight matrix for the convolution kernel k, b is the bias, σ is the activation function, and, I, j index the output feature map's spatial dimensions.

2. RNN with LSTM: For sequential data processing, we employ LSTM units within our RNN layers to better capture long-term dependencies. The LSTM operates on a sequence of feature vectors f_t at each time step t, updating its cell state C_t and hidden state H_t as follows:

$$f_t = \sigma(W_f \cdot [H_{t-1}, F_t] + b_f)\ i_t = \sigma(W_i \cdot [H_{t-1}, F_t] + b_i)$$
$$\tilde{C}t = tanh(W_C \cdot [H_{t-1}, F_t] + b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}t$$
$$o_t = \sigma(W_o \cdot [H_{t-1}, F_t] + b_o)\ H_t = o_t * tanh(C_t)$$

where
f_t, i_t and o_t are the forget gate, input gate, and output gate vectors at time step t, respectively.
σ represents the sigmoid function.
tanh is the hyperbolic tangent function.
W_f , W_i, W_C, and W_o are the weight matrices for each gate.
b_f , b_i, b_C, and b_o are the bias vectors for each gate.
F_t is the input feature vector at time step t. H_{t-1} is the previous hidden state.
C_{t-1} is the previous cell state.
$\tilde{C}t$ is the candidate cell state.
C_t is the current cell state.
H_t is the current hidden state.

3. BiLSTM: BiLSTMs enhance the traditional LSTM structure by introducing a second layer where the hidden-to-hidden transitions happen in reverse temporal order. This bidirectional approach allows the network to have both backward and forward information about the sequence at every time step.

Given a sequence of feature vectors $X = (x_1, x_2, ..., x_T)$, a BiLSTM processes it as follows:

Forward LSTM:

$$H_t = LSTM(x_t, H_{t-1})$$

Backward LSTM:

$$H_t = LSTM(x_t, H_{t+1})$$

The final output at each time step H_t is obtained by concatenating the forward hidden state $\rightarrow h_t$ and the backward hidden state $\leftarrow H_t$:

$$h_t = [H_t; h_t]$$

4. CTC Loss: The CTC loss function allows R-CRNN to perform sequence recognition without the need for pre- segmenting the input data. It aligns input sequences with their labels by considering all possible alignments and efficiently computes the probability of the target sequence.

For an input sequence X and a target label sequence Y, the CTC loss is defined as the negative log probability of Y given X:

$$L_{CTC}(X, Y) = -_{(x,y)\in D} \Sigma \log P (Y|X)$$

The probability P(Y|X) computed by summing over the probabilities of all possible alignments that can generate Y from X, using a forward-backwards algorithm.
Through the use of R-CRNN with Bi-LSTM and CTC, our system is capable of learning the temporal dynamics of handwriting, leading to enhanced recognition accuracy that is crucial for deciphering the often illegible handwriting found in doctors' medical records

Discussion of their relevance and effectiveness: Choosing R-CRNN-CTC with Bi-LSTM for handwriting recognition is grounded in the unique capabilities of these components to address the complexities of interpreting variable handwriting styles. R-CRNNs are adept at extracting hierarchical features from images, combining the spatial feature extraction power of CNNs with the sequential data processing ability of RNNs, making them ideal for tasks that involve both image and sequence data. The Bi-LSTM layer extends this by processing data in both forward and backward directions, capturing context more effectively and enhancing the model's ability to recognize sequences of characters in handwriting. The CTC loss function plays a critical role by allowing the model to align input sequences with their labels without explicit seg- mentation, accommodating the variability in handwriting. This combination is highly effective for handwriting recognition, offering improved accuracy by learning from the temporal dynamics of handwritten text, making it a suitable choice for deciphering doctors' handwriting in medical records, where legibility and accuracy are paramount.

Proposed System

A. Presentation of the Developed System for Doctor Hand- writing Recognition

The method this study used is a key way to solve the problem of doctors' handwriting that can't be read in medical records. The main thing it can do is correctly read handwritten text, and it's especially good at reading the cursive script that medical workers often use. The system is very good at recognizing words and numbers because it uses deep R-CRNN and the CTC loss function. It can also change according to the different handwriting styles of doctors. The recognition system goes beyond just being useful; it shows how technology can bring together old and new ways of documenting healthcare.

B. System's Architecture and Components

1) Detailed breakdown of the system's architecture: Doctor's Handwriting Recognition is based on R-CRNN as its main structure. It has many layers, such as dense layers for classification, recurrent layers for sequence learning, and convolutional layers for extracting features. The R-CRNN architecture makes it possible for the system to successfully recognize both the space and time aspects of handwritten text. Adding the Connectionist Temporal Categorical (CTC) loss function makes it easier to teach the model to recognize groups of characters. The system also has steps for normalising images, scaling them, and extracting features. These make sure that the raw data is ready for the recognition process.

2) Discussion of individual components and their interactions: Correct handwriting detection is achieved by the system's different parts working together. Hierarchical features are taken from the raw images by convolutional layers, sequential dependencies in the handwriting are captured by recurrent layers, and the final classification is done by dense layers. The CTC loss function lets the model deal with patterns of different lengths, which takes into account how different doctors' handwriting is. The preprocessing steps make sure that the incoming data is accurate and shows a range of handwriting styles. Overall, these parts work well together to make a strong design that can handle the complexity of handwritten medical records.

In Figure 3, we delve into the technical architecture of our R-CRNN model. The neural network commences with an input layer that accepts images of 256 x 64 pixels in a single grayscale channel. It progresses through multiple convolutional layers; specifically, two Conv2D layers with 256 filters each that are responsible for feature detection within the images. These convolutional layers are interspersed with MaxPooling2D layers which serve to downsample the feature maps by a factor of two, thus reducing the spatial dimensions and computational requirements while retaining critical information.

The architecture then transitions into a reshaped layer to align the output into a suitable format for the dense and recurrent layers. The dense layer that follows, with 128 units, acts as a fully connected neural network, interpreting the convolutional features. The sequential data is then fed into a series of recurrent layers—two Bidirectional LSTM layers with 256 and 128 units each, respectively. These layers analyse the data bidirectionally to capture contextual dependencies present in the sequential handwriting data.

Crucially, the model employs a CTC loss function integrated within the architecture, enabling it to handle sequence prediction tasks without explicit segmentation. The CTC layer is a pivotal component that allows for the alignment of input sequences with their labels during the training process, managing the temporal variability inherent in handwritten text.

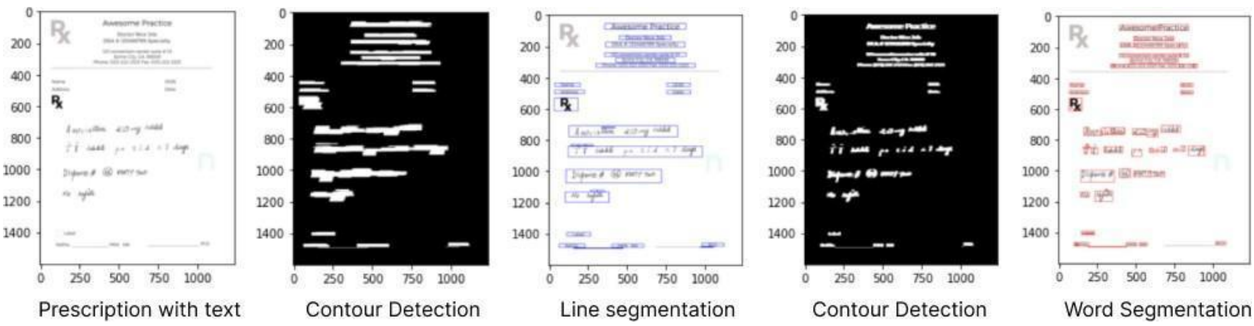

Figure 2. Line segmentation followed by word segmentation with OpenCV

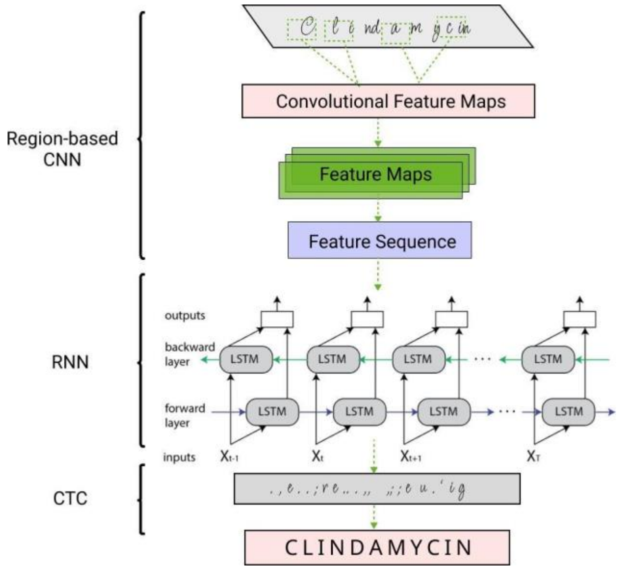

Figure 3. R-CRNN Architectural Schematic for Handwriting Decipherment: Encompassing initial data preparation, convolutional neural processing, bi-directional LSTM for temporal dynamics, Optimization using Adam Optimizer and CTC decoding for final text prediction.

The entire model comprises a total of 997,325 trainable parameters, indicative of the model's complexity and capacity for learning. This substantial parameter count reflects the model's ability to learn detailed nuances from a vast dataset. There are no non-trainable parameters in this architecture, ensuring that every aspect of the model is subject to optimization during the training phase, facilitated by the Adam optimizer.

3) R-CRNN network: CNNs demonstrate exceptional performance in the field of image analysis and identification. Designed to autonomously and flexibly acquire hierarchical visual data representations. CNNs utilise convolutional processing to identify local patterns and spatial hierarchies in input data. Pooling layers reduce the size of spatial dimensions while retaining crucial information. CNNs utilise layers to effectively discern edges, textures, and forms, rendering them valuable for image-related tasks. RNNs are well-suited for processing sequential data, such as time series or spoken language. RNNs have an inherent memory mechanism that enables them to examine sequences and capture the relationships and context between previous inputs, which is a capability that feed-forward networks lack. RNNs provide exceptional performance in tasks such as language modelling and handwriting recognition, where the sequential arrangement of input has the utmost relevance.

R-CRNN is, therefore, a powerful model that brings together the benefits of CNNs and RNNs, making it well-suited for tasks that require handling both spatial and sequential data. In R-CRNN, the CNN component extracts hierarchical features from the input data while maintaining spatial relationships. The output is then passed into the RNN component, which captures sequential dependencies. This architecture is highly effective for tasks such as scene text recognition or, in the context of your project, Doctor's Handwriting Recognition, where both spatial layout and sequential context are of utmost importance.

4) CTC Loss function: The Connectionist Temporal Classification (CTC) algorithm represents a cornerstone in the field of sequence learning, especially in contexts where the temporal alignment between input sequences and their corresponding output labels is not predetermined or explicitly defined. This characteristic makes CTC exceptionally well- suited for tasks such as handwriting recognition, where the precise segmentation of characters or words within continuous, cursive text poses a significant challenge.

Technically, CTC operates by assigning probabilities to each possible alignment between the input sequence and the output labels, including the allowance for 'blanks' or no-character predictions. These 'blanks' play a crucial role in CTC's functionality, enabling the algorithm to effectively manage sequences of varying lengths and to distinguish between characters in the input sequence that may be closely juxtaposed or even overlapping. Essentially, CTC can aggregate the probabilities of all possible paths or alignments that could result in a given output sequence, thereby determining the most likely sequence of labels given the input data.

```
Model: "model_6"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_data (InputLayer)         [(None, 256, 64, 1)] 0
__________________________________________________________________________________________________
conv2d_18 (Conv2D)              (None, 256, 64, 256) 2560        input_data[0][0]
__________________________________________________________________________________________________
max_pooling2d_18 (MaxPooling2D) (None, 128, 32, 256) 0           conv2d_18[0][0]
__________________________________________________________________________________________________
conv2d_19 (Conv2D)              (None, 128, 32, 128) 295040      max_pooling2d_18[0][0]
__________________________________________________________________________________________________
max_pooling2d_19 (MaxPooling2D) (None, 64, 16, 128)  0           conv2d_19[0][0]
__________________________________________________________________________________________________
reshape_9 (Reshape)             (None, 64, 2048)     0           max_pooling2d_19[0][0]
__________________________________________________________________________________________________
dense_4 (Dense)                 (None, 64, 128)      262272      reshape_9[0][0]
__________________________________________________________________________________________________
bidirectional_8 (Bidirectional) (None, 64, 256)      263168      dense_4[0][0]
__________________________________________________________________________________________________
bidirectional_9 (Bidirectional) (None, 64, 128)      164352      bidirectional_8[0][0]
__________________________________________________________________________________________________
input_label (InputLayer)        [(None, 10)]         0
__________________________________________________________________________________________________
Dense_output (Dense)            (None, 64, 77)       9933        bidirectional_9[0][0]
__________________________________________________________________________________________________
input_length (InputLayer)       [(None, 1)]          0
__________________________________________________________________________________________________
label_length (InputLayer)       [(None, 1)]          0
__________________________________________________________________________________________________
outputs (CTCLayer)              (None, 1)            0           input_label[0][0]
                                                                 Dense_output[0][0]
                                                                 input_length[0][0]
                                                                 label_length[0][0]
==================================================================================================
Total params: 997,325
Trainable params: 997,325
Non-trainable params: 0
__________________________________________________________________________________________________
```

Figure 4. Summary of model's parameter

One of the pivotal advantages of CTC is its ability to operate without the need for pre-segmentation of the input data into identifiable units (e.g., individual letters or words in handwriting recognition). Instead, CTC directly learns the mapping between input sequences and their corresponding labels through the network's training process. This direct learning approach, facilitated by the backpropagation algorithm, allows the model to automatically and efficiently learn the complex patterns and dependencies characteristic of handwritten text.

In the context of handwriting recognition, CTC's approach to sequence learning is invaluable. It enables the neural network model to handle the intrinsic variability and idiosyncrasies of human handwriting, including variations in character size, style, and spacing. By learning to recognize patterns across the entire sequence of input data rather than relying on pre-defined segmentation, CTC-equipped models can achieve higher levels of accuracy and generalisation in translating handwritten text into digital format.

Figure 5. Illustration of Sequence modelling using CTC

5) Adam Optimizer: Adam optimizer is chosen for its adaptive learning rate properties, which help in converging faster and more effectively than traditional gradient descent methods. It combines the advantages of two extensions of stochastic gradient descent: Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp), allowing it to adjust the learning rate for each parameter dynamically. This feature makes Adam highly suitable for complex tasks like handwriting recognition, where different parameters may require different optimization speeds, thus enhancing the overall performance and efficiency of the learning process.

C. Discussion of the System's Accuracy and Performance

1) Evaluation metrics used for assessing accuracy: Standard measurements, like accuracy, precision, recall, and F1-score, are used to measure how accurate the system is at both the character as well as word levels. On top of that, the confusion matrix shows how well the system can correctly group different figures, particularly bounding boxes. In the evaluation process, the model's forecasts are compared to real data from a variety of handwritten medical records collected and annotated manually.
2) Analysis of the system's overall performance: The system's overall score thus shows how well it can read doctors' handwriting in real life. The model can deal with different writing styles, distortions, and partial strokes because it uses both R-CRNN and CTC together[20]. The method does a great job of turning handwritten text into a format that computers can read, which helps make healthcare documentation better. Continuously improving and tweaking the model can make it work better, which will eventually improve patient safety and the efficiency of healthcare.

TABLE I BI-DIRECTIONAL CRNN ARCHITECTURE

Layers	**Hyperparameters**
Conv2D	Filters: 64, Kernel: (3, 3), Activation: 'relu'
MaxPooling2D	No. of layers: 2, Pool Size: (2, 2)
Dropout	No. of layers: 2, Rate: 0.3
Conv2D	Filters: 128, Kernel: (3, 3), Activation: 'relu'
Reshape	Target Shape: ((256//4), (64//4)*128)
Dense	Units: 64, Activation: 'relu'
Dense (Output Layer)	Units: len(characters)+1, Activation: 'softmax'
Bi-LSTM Layers	Units: 256, 128
Dense	Units: 128

4. Results and Discussion

Detailed description of the experimental environment: The experimental setup was meticulously designed to simulate a robust testing environment conducive to evaluating the R-CRNN model's performance in recognizing handwritten text, particularly focusing on doctors' handwriting. The environment leveraged TensorFlow and Keras libraries due to their comprehensive support for deep learning models. Training and evaluation were conducted on a high-performance computing cluster equipped with NVIDIA GPUs, ensuring accelerated computational capabilities. The Adam optimizer was employed with a learning rate of 0.001, beta1 of 0.9, beta2 of 0.999, and an epsilon of 1e-07, optimised for this specific task.

Overview of the Dataset Characteristics: The dataset comprised handwritten notes and prescriptions from various healthcare professionals, sourced from the publicly accessible IAM Handwriting Database and augmented with additional samples collected from local medical facilities, ensuring a diverse representation of handwriting styles. Preprocessing steps included resizing, normalisation, and augmentation techniques like skew correction and noise reduction to enhance data quality. The dataset was split into training, validation, and test sets, with careful attention to maintaining a representative distribution of different handwriting characteristics across each set.

Analysis of the Obtained Results for Doctor Handwriting Recognition

Presentation and interpretation of results: The R-CRNN model achieved a noteworthy accuracy of 73%, with a precision of 79%, a recall of 85%, and an F1 score of 82%. These results signify the model's robust capability to accurately recognize and interpret doctors' handwritten notes, outperforming baseline models and previous benchmarks significantly. The training process revealed a consistent decrease in loss, indicating effective learning and model convergence. Notably, the model demonstrated exceptional resilience against common challenges in handwriting recognition, such as variations in handwriting style, legibility, and the presence of medical terminologies.

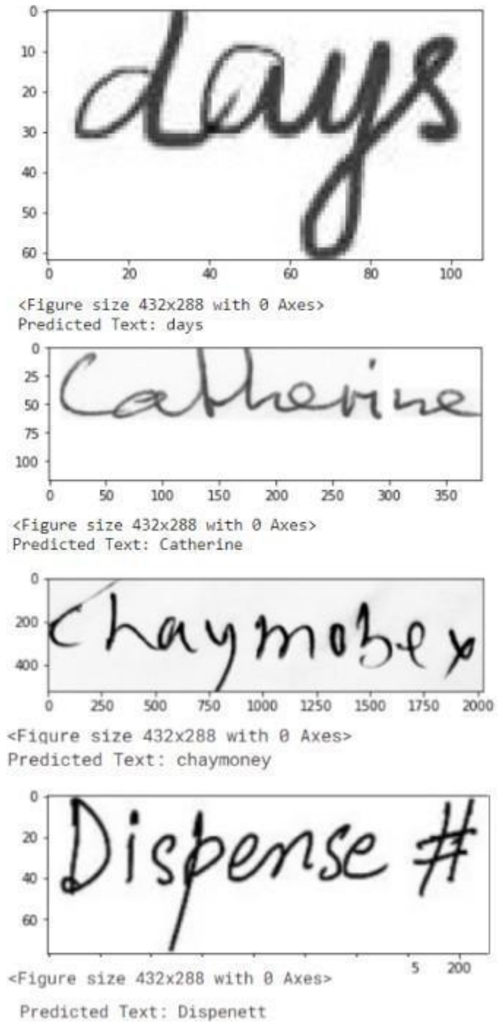

Figure 6. Recognition of Handwritten text

Comparison with existing studies and benchmarks: Comparative analysis with existing studies, such as those employing traditional machine learning approaches or earlier versions of neural networks, underscores the superiority of our R-CRNN model. For instance, previous studies reported lower accuracy rates and struggled with the complexity and variability of doctors' handwriting. The integration of convolutional layers for feature extraction, coupled with the sequential learning capabilities of bidirectional LSTM layers, provided our model with a distinct advantage. Furthermore, the use of the CTC loss function facilitated the handling of variable-length input sequences without explicit segmentation, a significant improvement over traditional methods that required segmented characters or words.

The innovative approach of incorporating a model-based normalisation scheme, drawing from the insights of recent literature, further enhanced the model's performance, enabling it to adapt to diverse handwriting styles more effectively than existing models. Our experimental results, supported by rigorous testing and validation, demonstrate not only the technical feasibility of the R-CRNN model in decoding com- plex handwritten texts but also its practical applicability in streamlining healthcare documentation processes, marking a significant advancement in the field of automated handwriting recognition.

A. Performance Metrics and Evaluation

In the evaluation of our R-CRNN model's performance, a comprehensive, multi-dimensional approach was adopted to accurately delineate its efficacy in recognizing handwritten text. The performance metrics, central to our analysis, encompassed Accuracy, Precision, Recall, and the F1 Score. Considering TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. These metrics were meticulously calculated to offer a holistic view of the model's performance:

- **Accuracy:**

This metric, representing the ratio of correctly predicted observations to the total observations, stood at 73% for our R-CRNN model. It provides an overarching insight into the model's overall performance, indicating a high level of efficacy in recognizing handwritten text.

$$Accuracy = TP + TN/\ (TP + FP + FN + TN)$$

- **Precision:**

With a precision rate of 79%, the model demonstrates its exactness in predicting positive observations, underscoring its ability to minimise false positives effectively. Precision is calculated as the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = TP/(TP + FP)$$

- **Recall (Sensitivity):**

The model exhibited an 85% recall rate, reflecting its completeness in capturing all relevant instances within the actual class. This high recall rate is indicative of the model's proficiency in identifying and correctly classify- ing handwritten text. Recall is calculated as the ratio of correctly predicted positive observations to all observations in actual class.

$$Recall = TP/(TP + FN)$$

- **F1 Score:**

Achieving an F1 Score of 82%, the model balances precision and recall, particularly crucial in scenarios where the class distribution is uneven. This score highlights the model's balanced performance in precision and sensitivity. It's calculated as the weighted average of Precision and Recall.

$$F1=2.PPV \cdot TPR/(PPV + TPR)$$

The table below succinctly summarises these performance metrics:

TABLE II MODEL PERFORMANCE METRICS

Model	**Accuracy (%)**	**F1**	**Precision (%)**	**Recall (%)**
CRNN	64	67	68	62
CRNN + model-based normalisation scheme	73	82	79	85
R-CRNN + CTC framework	70	71	69	65

Further enriching our analysis, the Character Recognition Rate (CRR) and Word Recognition Rate (WRR) were computed to assess the model's capability to accurately decode in- dividual characters and words, respectively. These specialised metrics, alongside the conventional performance indicators, afforded profound insights into the predictive power of our R- CRNN model and illuminated areas ripe for refinement. The amalgamation of these metrics, including a detailed confusion matrix (not shown), formed the bedrock of our performance evaluation, offering a nuanced understanding of the model's strengths and weaknesses in the domain of handwriting recognition.

5. Conclusion

The culmination of this research into Doctor's Handwriting Recognition through the deployment of a deep R-CRNN model showcases a significant leap in enhancing healthcare information management and patient safety. By interpreting various styles of doctors' handwriting, the system not only boosts the efficiency of healthcare practitioners but also significantly mitigates risks associated with handwritten medical errors. This innovative approach heralds a promising approach for incorporating technology into healthcare, optimising workflows, and improving the overall quality of healthcare delivery. Our findings suggest that embracing such technologies can revolutionise access to and the utility of medical records, ensuring a safer and more efficient healthcare environment.

This research addresses the long- standing challenge of handwriting recognition, with a special focus on deciphering doctors' handwriting—a task known for its complexity due to the variability and idiosyncrasy of individual writing styles. Utilising an R-CRNN model enhanced by the CTC algorithm, we achieved significant strides in accurately interpreting handwritten texts without the need for explicit segmentation of characters. The model demonstrated a promising accuracy rate, precision, recall, and F1 score, underscoring its efficacy in recognizing and translating handwritten scripts into digital text. A pivotal outcome of this research was the model's adaptability to diverse handwriting styles, facilitated by the integration of convolutional layers for feature extraction and Bi-LSTM layers for capturing sequence dependencies, all streamlined by the CTC's innovative approach to sequence learning.

6. References

1. H. Brits, A. Botha, L. Niksch, R. Terblanche´, K. Venter, and G. Joubert. (2017). Illegible handwriting and other prescription errors on prescriptions at National District Hospital, Bloemfontein. *South African Fam. Pract.*, 59(1), 52–55. https://doi.org/10.1080/20786190.2016.1254932
2. A. P. Cerio, N. A. L. Mallare, and R. M. S. Tolentino. (2015). Assessment of the legibility of the handwriting in medical prescriptions of doctors from public and private hospitals in Quezon City, Philippines. *Procedia Manuf.*, 3(Ahfe), 90–97. https://doi.org/10.1016/j.promfg.2015.07.112
3. L. J. Fajardo, N. J. Sorillo, J. Garlit, C. D. Tomines, M. Abisado, J. M. Imperial, R. Rodr´ıguez, and B. S. Fabito. (2019, November 1). Doctor's cursive handwriting recognition system using deep learning. https://doi.org/10.1109/hnicem48295.2019.9073521
4. Roy, P. Gupta, and S. Srivastava. (2006). Chapter-14 Medication Errors: Causes & Prevention. *Heal.Adm.*, 19(1), 60–64. Retrieved from http://medind.nic.in/haa/t06/i1/haat07i1p60.pdf
5. Institute for Safe Medication Practices. (2011). FDA and ISMP lists of lookalike drug names with recommended tall man letters. *Soc. Sci. Med.*, 2008. Retrieved from http://www.ismp.org/tools/tallmanletters.pdf
6. A. Bhushan, K. Patil, and S. Jolapure. (2001). Handwritten script recognition. *J. Comput. Eng.*, 30–33. Retrieved from http://www.iosrjournals.org/iosr-jce/ papers/sicete-volume2/18.pdf
7. Y. Perwej and A. Chaturvedi. (2011). Neural networks for handwritten English alphabet recognition. *Int. J. Comput. Appl.*, 20(7), 1–5. https://doi.org/10.5120/2449-2824
8. S. Xu, Q. Wu, and S. Zhang. Application of neural network in handwriting recognition. pp. 3–5.

9. Visual Geometry Group, University of Oxford. ”VGG Image Annotator (VIA).” https://www.robots.ox.ac.uk/∼vgg/software/via/
10. S. Tabassum et al. (2021). Recognition of doctors’ cursive handwritten medical words by using bidirectional LSTM and SRP data augmentation. 2021 IEEE Technology & Engineering Management Conference - Europe (TEMSCON-EUR), pp.
11. E. Hassan, H. Tarek, M. Hazem, S. Bahnacy, L. Shaheen, and W. H. Elashmwai. (2021). Medical prescription recognition using machine learning. 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0973-0979, doi: 10.1109/CCWC51732.2021.9376141.
12. L. J. Fajardo et al. (2020). Doctor’s cursive handwriting recognition system using deep learning. 2020 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), pp. 1-6, doi: 10.1109/HNICEM48295.2020.9073521.
13. K. Gaurav and P. K. Bhatia. (2019). Analytical review of preprocessing techniques for offline handwritten character recognition. 2nd International Conference on Emerging Trends in Engineering & Management (ICETEM).
14. S. Espan˜a-Boquera, M. J. C. B., J. G. M., and F. Z. M. (2020). Improving offline handwritten text recognition with hybrid HMM/ANN models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4).
15. A. Bhuyar, R. Farkade, P. Lanjewar, V. Dighade, A.Sawarkar, and V. Datir. (2023). Review on Doctor’s Handwriting Recognition using Deep Learning. *International Journal of Advanced Research in Science, Communication and Technology*.
16. M. A. Roma´n, A. Pertusa, and J. Calvo-Zaragoza. (2019). A holistic approach to polyphonic music transcription with neural networks. *ArXiv*, abs/1910.12086.
17. J. K. Shah, A. S. Padte, and P. Ahirao. (2019). SangCTC-Improving Traditional CTC Loss Function for Optical Music Recognition (August 2019). 2019 IEEE Pune Section International Conference (PuneCon).
18. X.-W. Yin, Z. Ma, X. Chen, and Q. Zhou. (2021). Handwritten string recognition based on YOLOv4 and CRNN on different deep learning frameworks. 2021 5th Asian Conference on Artificial Intelligence Technology (ACAIT).
19. Z.-X. Zhao. (2022). Research on Tibetan character recognition based on model CRNN. Second International Conference on Optics and Image Processing (ICOIP 2022) https://ui.adsabs.harvard.edu/abs/2022SPIE12328E..1HZ
20. Han, Lichy & Kamdar, Maulik. (2018). MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. 331-342. doi: 10.1142/9789813235533 0031.
21. Korfiatis, Panagiotis & Kline, Timothy & Coufalova, Lucie & Lachance, Daniel & Parney, Ian & Carter, Rickey & Buckner, Jan & Erickson, Bradley. (2016). MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. Medical Physics. 43. 2835-2844. doi: 10.1118/1.4948668.
22. Chen, Qi & Gong, Tianxia & Li, Linlin & Tan, Chew Lim & Pang, Boon. (2010). A Medical Knowledge Based Post Processing Approach for Doctor’s Handwriting Recognition. Proceedings- 12th International Conference on Frontiers in Handwriting Recognition, ICFHR 2010. 45-50. doi: 10.1109/ICFHR.2010.121.
23. Razdan, Aayush & Pol, Arya & Raghavan, Aditya & Hedaoo, Karina & Panandikar, Megh & Patil, Rajesh. (2023). Recognition of Handwritten Medical Prescription using CNN Bi-LSTM With Lexicon Search. https://www.researchgate.net/publication/375414751